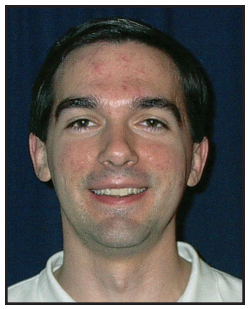
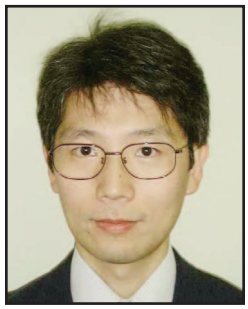
Session 4 Overview / High-Performance Digital Enterprise Processors & Components





Session Chair: *Joshua Friedrich, IBM Systems and Technology Group, Austin, TX*

Session Co-Chair: *Takashi Miyamori, Toshiba, Kawasaki, Japan*

Session 4 focuses on significant new innovations in the development of enterprise class microprocessors. This year's enterprise designs feature the fastest clock frequency, the highest x86 core count, and the highest energy efficiency, and the highest transistor count ever achieved in commercial microprocessors. The trend of dramatically increasing the number of threads, special function units, and off-die interconnect contained within a single die also continues. These designs from IBM, Intel, AMD, and the Chinese Academy of Sciences also employ a variety of power-efficient circuit and micro-architectural techniques to continue Moore's Law in the face of the dramatic power challenge plaguing deep submicron technology.

The first paper describes the microprocessor chip for the IBM zEnterprise 196 system, which is the first commercial processor chip to ship at a frequency in excess of 5 GHz. The 512mm2, 5.2GHz chip was built in IBM's 45nm Silicon-on-Insulator CMOS process with 13 levels of copper interconnect. It contains 4 processor cores, 1.5MB of private SRAM L2 cache per core, and an on-chip, high-speed 24MB shared L3 cache built from IBM's unique embedded DRAM technology that combines DRAM density with a high-speed digital logic process. To meet the incredibly high frequency objective, the design team made major cycle time enhancements, reduced core power consumption at constant PVT conditions by over 25% compared to the 65nm Z core, and improved the on-die power distribution. At the same time, the design added out-of-order instruction processing to further improve performance at constant frequency.

Paper 4.2 highlights a particularly crucial high speed component of IBM's zProcessor core – the dynamic hit logic and embedded 8Kbit SRAM used for L1 cache hit detection. The 14 bit hit logic described in this paper uses programmable launch and reset clocks along with a mix of dynamic and highly skewed static CMOS to achieve record operating frequencies (above 6.5GHz at lab conditions). In addition, its innovative use of a "search-for-a-hit" scheme minimizes power consumption without degrading performance. Array BIST provides both the hit logic and SRAM with full "at-speed" test coverage for the first time.

Paper 4.3 from Intel's lab in Bangalore, India, describes the next generation enterprise Xeon® processor consisting of 10 Westmere 32nm cores and a shared, inclusive L3 cache integrated on a monolithic die with link based IOs. The die contains the highest number of x86 cores ever integrated onto a single chip and contains significant innovations and circuit optimizations over its predecessor, Nehalem, targeting idle power reduction, robust high speed IO links at the next generation process node, and performance per watt improvements. The processor is implemented in 32nm CMOS using high-κ metal gate transistors and nine copper interconnect layers. It supports individually controlled power gating of each of the ten cores and implements macro clock gating of uncore and cache functions to drive idle power to new lows.

Energy efficient processing is also a key area of innovation, and the Godson-3B processor from the Chinese Academy of Sciences, described in Paper 4.4, reports the highest energy efficiency (3.2GFlops/Watt) among all state-of-art high-performance processors. Godson-3B is an 8-core design implemented on 65nm CMOS LP/GP mixed process with 7 layers of Cu metallization. It contains 582.6M transistors within 299.8mm2 area and operates at frequencies of up to 1.05GHz. Its peak performance is 128/256GFlops for double/single-precision with 40W power consumption.

Paper 4.5 from AMD describes the new x86-64 based 2-core CPU module (Bulldozer) that contains 213M transistors in 11-metal layer 32nm high-k metal-gates SOI CMOS process. In addition to the micro-architecture improvements, the components, such as the L1 and L2 caches, the integer unit and the Floating Point unit, are designed to achieve higher frequency, lower power consumption, and lower gate counts per cycle than the 45nm AMD core while maintaining IPC (Instructions per Cycles). It operates over 3.5GHz in an area (including 2MB L2 cache) of 30.9mm2.

The next paper 4.6 from AMD presents details of the 40-entry unified, out-of-order scheduler and integer unit of the Bulldozer. The scheduler issues four operations per cycle and the integer execution unit supports single-cycle bypass between four functional units. Instead of dynamic logic, skewed gates with static standard cell are extensively used. Furthermore, to minimize power consumption, the design reduces switching activity by manipulating pointers rather than moving large amounts of data within queues.

 978-1-61284-302-5/11/$26.00

Paper 4.7 from Intel describes the clock generation system of a multi-core processor for server applications, fabricated on a 9-metal layer 32nm CMOS process. A growing number of CPU cores and integration of high speed I/O ports, such as QuickPath Interconnect, PCI express and DDR3, presents significant challenges to the clock generation and distribution systems. The proposed clock system architecture delivers low skew, low power and low latency with modularity and scalability for a ring-like architecture in spite of these challenges.

Finally, Paper 4.8 from Intel is the Itanium processor implemented in a 32nm CMOS process with 9 layers of copper interconnection. It contains 3.1 billion transistors within a 18.2 mm by 29.9 mm die. This transistor count represents a 50% increase over the previous Itanium processor and is the highest transistor count ever reported for a single die. The processor includes eight multi-threaded cores, a ring based system interface, and more than 54MB of cache memory. A three level cache hierarchy consists of a first level single cycle 16K Instruction (I) and Data (D) cache that is backed by two second level caches (a nine cycle 512K I cache and an eight cycle 256K D cache) and a 32MB the last level cache. The design implements twice as many as cores as the previous Itanium while lowering the thermal design power (TDP) by 15W to 170W. High speed links allow for peak processor-to-processor bandwidth of up to 128 GB/s and memory bandwidth of up to 45GB/s.

Together, these innovative enterprise processors and design components represent a new height of performance, integration, and energy efficiency. By continuing to deliver exponentially increasing performance and capability, these efficient engines will enable computers to analyze our world's most challenging scientific problems and manage the needs of increasingly complex global enterprise.

Presenters:

1:30 PM

4.1 A 5.2GHz Microprocessor Chip for the IBM zEnterprise™ System

J. Warnock, IBM Systems and Technology Group, Yorktown Heights, NY

2:00 PM

4.2 Dynamic Hit Logic with Embedded 8Kb SRAM in 45nm SOI for the zEnterprise™ Processor

A. R. Pelella, IBM Systems and Technology Group, Poughkeepsie, NY

2:15 PM

4.3 A 32nm Westmere-EX Xeon® Enterprise Processor

S. Sawant, Intel, Bangalore, India

2:30 PM

4.4 Godson-3B: A 1GHz 40W 8-Core 128GFLOPS Processor in 65nm CMOS

W. Hu, Chinese Academy of Sciences, Beijing, China; Loongson Technologies, Beijing, China

3:15 PM

4.5 Design Solutions for the Bulldozer 32nm SOI 2-Core Processor Module in an 8-Core CPU

H. McIntyre, AMD, Sunnyvale, CA

3:45 PM

4.6 40-Entry Unified Out-of-Order Scheduler and Integer Execution Unit for the AMD Bulldozer x86-64 Core

M. Golden, AMD, Sunnyvale, CA

4:15 PM

4.7 Clock Generation for a 32nm Server Processor with Scalable Cores

S. Li, Intel, Santa Clara, CA

4:45 PM

4.8 A 32nm 3.1 Billion Transistor 12-Wide-Issue Itanium® Processor for Mission-Critical Servers

R. J. Riedlinger, Intel, Fort Collins, CO